NAVAL POSTGRADUATE SCHOOL
MONTEREY, CALIFORNIA

THESIS

USER INTERFACE OPTIMIZATION THROUGH BREADTH OF DISTRIBUTION ANALYSIS

by

Luis Enrique Reinoso

March, 1997

Thesis Co-Advisors: Timothy Shimeall
John Daley

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE March 1997	3. REPORT TYPE AND DATES COVERED Master's Thesis

4. TITLE AND SUBTITLE USER INTERFACE OPTIMIZATION THROUGH BREADTH OF DISTRIBUTION ANALYSIS	5. FUNDING NUMBERS
6. AUTHOR(S) Reinoso, Luis E.	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey CA 93943-5000	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

12a. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution is unlimited.	12b. DISTRIBUTION CODE

13. ABSTRACT *(maximum 200 words)*

A Java tutorial was developed as a World Wide Web (WWW) site for use in capturing user behavior data. Breadth of distribution analysis was then applied to the data collected in order to characterize the usage of the user interface through the shape, connectedness, and order of traversal of each user in the sample.

The results reveal distinct user groups with different levels of user knowledge and needs in relation to the web site content. The resulting user interface analysis process produces a set of recommendations for optimizing the user interface, including adaptive interfaces for different user sub-groups and optimization of sequential rather than topical presentations.

14. SUBJECT TERMS optimization, distribution, user interface analysis	15. NUMBER OF PAGES 73
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT UL

Approved for public release; distribution is unlimited.

USER INTERFACE OPTIMIZATION THROUGH BREADTH OF DISTRIBUTION ANALYSIS

Luis Enrique Reinoso
Lieutenant, United States Navy
B.S., Northwestern University, 1990

Submitted in partial fulfillment
of the requirements for the degree of

MASTER OF SCIENCE IN COMPUTER SCIENCE

from the

NAVAL POSTGRADUATE SCHOOL
March 1997

Author: ______________________________
Luis Enrique Reinoso

Approved by: ______________________________
Timothy Shimeall, Thesis Co-Advisor

____
John Daley, Thesis Co-Advisor

____
Theodore Lewis, Chairman
Computer Science Department

ABSTRACT

A Java tutorial was developed as a World Wide Web (WWW) site for use in capturing user behavior data. Breadth of distribution analysis was then applied to the data collected in order to characterize the usage of the user interface through the shape, connectedness, and order of traversal of each user in the sample.

The results reveal distinct user groups with different levels of user knowledge and needs in relation to the web site content. The resulting user interface analysis process produces a set of recommendations for optimizing the user interface, including adaptive interfaces for different user sub-groups and optimization of sequential rather than topical presentations.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

The author wishes to acknowledge the support of Elliotte Rusty Harold, for allowing the use of his Java tutorial as a testbed for this thesis.

The author also would like to thank the students of the Computer Science, Electrical Engineering, Information Technology Management, and Information Warfare curricula for their help in using the testbed for this thesis.

The author finally would like to thank Prof. Timothy Shimeall and LCDR John Daley for their guidance and patience during the work in performing this study.

I. INTRODUCTION

A. PREFACE

An interface enables certain operator behaviors, is neutral toward other behaviors, and inhibits or prohibits still other behaviors (Bass, 1991). As a result of rapid advances in modern technology, the amount and complexity of the information provided for the user to assimilate in a short period of time increases the likelihood of inefficient, incorrect, and at worst unsafe use of the interface (Rencken, 1993). The reduction of cognitive overhead and disorientation in a particular interface is essential for the continued usefulness of the interface (Balasubramanian, 1995).

Several user interface design methodologies exist with the goal in mind of providing the user a useful and appropriate interface. The four primary views on human computer interaction (HCI) are:

- interaction-oriented: measures usability quality in terms of how the user interacts with the interface;

- user-oriented: defines usability quality in terms of the mental effort and attitude of the user;

- product-oriented: defines usability quality in terms of ergonomic attributes of the interface itself;

- formal: usability is formalized and simulated in terms of mental models.

Each view has its set of measurements, with which the inclusion of breadth of distribution analysis data can yield useful information towards usability recommendations (Rauterberg, 1996).

User interfaces are commonly evaluated using centrally-based measures that may ignore large populations of "non-average" users. These centrally-based measures include the use of mean and median. For example, in a study by Rauterberg on user interface characteristics, the measures for quantifying functional feedback, interactive directness, application flexibility and dialog flexibility use averages for their comparative usability studies (Rauterberg, 1996). Catledge and Pitkow use average slope to characterize user navigation strategies in hypertext browsing and searching (Catledge, 1995).

Breadth of distribution analysis is the use of measures of extent (e.g., Variance, standard deviation, range, quartile, etc.) of a distribution as the basis for analyzing a user interface. Breadth of distribution analysis can quantify the degree of variance in optimization for widely heterogeneous groups of users, and can offer insight into multi-level user interfaces that take into account different needs among a largely disparate user base. By examining the spread of values over a particular graph, breadth of distribution analysis breaks out the sub-groups that may exist within a user population as represented pictorially by separate but distinct curves in the graph. Thus, breadth of distribution analysis provides a method for distinguishing, for example, two graphs that have similar values for mean and median, but differ visibly in their graphical representation.

After selecting and designing a prototype, statistical methods in determining the breadth of distribution are implemented to characterize the usage of the user interface. By collecting information on actual user behaviors, concrete data is compiled on usage of different areas of the user interface. From these data, a set of recommendations can be

made for possible refinements to the interface in order to more closely match the needs of the user population.

B. IMPORTANCE OF USER INTERFACE ANALYSIS

With the continued growth of information technologies including the World Wide Web (WWW), the accessibility of these technologies has expanded to include users of diverse levels of experience and skill (Pitkow, 1995). Physicians, lawyers, corporate executives, teachers and students at various levels of schooling have differing levels of computer literacy, frequency of computer use, and requirements for as well as interests in information technologies (Pitkow, 1995). The variety of all these factors reflects a population diversity that will continue to expand with the growth of these technologies.

In a combat situation, an interface that fails to meet the requirements of the user may result in disaster. For example, on July 3, 1988, an ascending Iran Air A300 Airbus over the Persian Gulf was interpreted as a descending hostile aircraft by the air search console operator and was shot down by USS Vincennes. Other incidents such as the Three Mile Island nuclear accident illustrate the importance

of an interface that recognizes user needs and limitations, thus minimizing the potential for human error (Malone, 1990).

The enduring success of information technologies is dependent on their ability to adjust to the disparate needs and interests of the user population. In addition to understanding the diversity of the user population backgrounds, detailed information on the actual use of these technologies is necessary (Catledge, 1995). With the continued growth in accessibility of these technologies, they hold the potential of ushering in the "age of information" to people of all ages, backgrounds, and economic status (Pitkow, 1995).

The development of intelligent user-centered interfaces provides for intuitive HCI, which in turn contributes to greater initial and long-term interest in information technology. Computer software and hardware continue to expand our potential to complete tasks more quickly and effectively than ever before. Yet a user's ability to make use of computers is limited by the degree of consistency and intuitiveness of the interface encountered (Dumas, 1988). Many less-experienced computer users who question the usefulness of computers in completing daily tasks do so

because of negative first experiences with cumbersome, inconsistent, and frustrating interfaces.

Thus, a keen understanding of the present and future user population is necessary for the design of user interfaces responsive to user needs and interests. Although the field of user interfaces is in its nascent stages, some of the components of an ultimate environment for user interface development are on the distant horizon (Bass, 1991). In order to achieve an understanding of the levels of computer literacy in the user population, the interface developers must see the interface as a vehicle for communication with the user (Dumas, 1988). Since the user interface can account for nearly 50 percent of total life cycle costs for interactive systems, the software engineer has a vested interest in designing a user interface that both satisfies the needs of the customer and is constructed using the best available tools and techniques available (Bass, 1991).

C. DESCRIPTION OF METHODS

One method for assessing user preferences is through surveys. Surveys provide a means of collecting data on user

responses to a set of questions. The relevance of the data gathered depends heavily upon the design of the survey so that all aspects of a user's preferences related to the interface are addressed. The use of surveys in conjunction with other tools, such as log file analysis, enable the development of effective user-centered interfaces (Pitkow, 1995).

Log file analysis enables the collection of concrete data on actual user behaviors associated with a particular user interface (Catledge, 1995). This method produces information on user visits and growth, activity levels, specific geographic and organizational origins of the user population, downloads and usage patterns, and time spent on particular areas of the interface. By examining the results of a log file analysis, conclusions can be made with regards to user interest in certain topic areas within the interface, as well as areas of difficulty in usage.

D. PROBLEM STATEMENT

How can statistical hypothesis testing on user interface usage be used as a basis for optimization of the interface? The complex and highly creative nature of user

interface design combines task analysis, an understanding of user requirements, experience, cognitive models, and intuition in creating a usable user interface (Balasubramanian, 1995). Hypothesis testing (or significance testing) is employed in user interface design to test some assumption (hypothesis) we have about the targeted user population against a sample from the population. The result of a significance test is a probability that we attach to a descriptive statistic calculated from a sample. Hypothesis testing allows us to evaluate differences between what we observe and what we expect on the basis of our hypothesis, but only in terms of the probability that these differences could have occurred by "chance" (Henckel, 1976).

How is significance defined in the absence of rigorous specifications for the interface goals? The use of usability standards with regards to software quality fails to measure all relevant product features in a task independent way (Rauterberg, 1996). Statistical inference allows us to characterize a population on the basis of a sample, and state how good this characterization is, or decide how well that sample fits our preconceptions of the population from which we drew the sample (Henckel, 1976).

E. DESCRIPTION OF THESIS

A survey of current literature shows several aspects inherent to the process of quantifying usability analysis. Numerous studies explore the general issue of measuring user interface characteristics towards the goal of improved usability, with the emphasis on central tendency figures. These studies also address the problem of selecting measurement criteria in the face of various models for user interface design.

After presenting the background on current applicable research in usability testing, an examination of various statistical methods determined the applicability of these methods for use in breadth of distribution analysis. The application of the selected statistical provided a means for maximizing the relevance of the data collected on user behaviors. The selection of specific statistical methods for use in breadth of distribution analysis facilitated the selection and design of a prototype for hypothesis testing.

In selecting a model for use in applying breadth of distribution analysis towards usability measurements, an interface for which a potential user population exists had to be selected and designed. The approach taken to solve this problem was to build a World Wide Web (WWW) site for

the purpose of data capture on user behaviors in accessing the site. The content of the site was selected based on an examination of potential need within the user population.

After selecting and designing a prototype, the implementation of the selected statistical methods provided the ability to characterize the usage of the user interface through breadth of distribution analysis. Once information is collected on usage patterns from the user sample, concrete data was compiled on the various interaction points of the user interface. From these data, a set of recommendations were made for possible refinements to the interface in order to more closely match the needs of the anticipated user population.

II. PREVIOUS WORK

A. PREFACE

Several studies examine the goal of improved usability by quantifying user interface characteristics through the use of central tendency figures, while addressing the issue of selecting measurement criteria in the face of various user interface design models. These studies explore various types of user interfaces, and in each case specific measurement criteria are defined in quantifying the user interface characteristics in order to create a set of recommendations for improved usability. Breadth of distribution analysis can further illuminate the characteristics of a user population, resulting in a set of recommendations for addressing the needs of a largely disparate user base.

B. ADAPTIVE TASK ALLOCATION IN HUMAN-COMPUTER INTERFACES

Rencken and Durrant-Whyte design and implement an HCI adaptive task allocation that is capable of on-line adjustment of the user's workload according to his/her ability to perform tasks for systems where the computer acts as a backup decision maker. By measuring service and error rates, the user's ability to accomplish tasks is continually evaluated to determine whether or not the user is capable of handling the next task, whereby the computer resolves the need to aid the user. Since the task allocation in the interface was divided into separate modules, which in turn incorporated quantification methods specific to each module, a qualitative analysis was used to assess the behavior of the user interface (Rencken, 1993). The results from this study illustrate the uniqueness of every user with regards to a particular interface, which in turn necessitates an interface capable of meeting the needs of each individual user.

C. USER INTERFACE DESIGN FOR HYPERTEXT SYSTEMS

Balasubramanian proposes a systematic framework for the design of user interfaces for hypertext systems based on Guilford's Theory of Intellect Model, wherein links and nodes are both considered typed objects. Such an approach proposes to help designers understand user requirements and the task domain, and significantly minimize disorientation, ambiguity, and cognitive overhead, by explicitly capturing the full range of semantic relationships between objects in the interface domain. A notable aspect of the proposed hypertext system design guidelines is that they allow non-sequential implementation, as the author views the user interface design process as non-linear and creative. An application of the principles addressed in this study aided the initial design of a WWW site model for use in applying breadth of distribution analysis towards usability measurements.

D. MAXIMUM LIKELIHOOD SENSITIVITY ANALYSIS AND NUISANCE PARAMETER ANALYSIS

In examining the connection between maximum likelihood sensitivity analysis and nuisance parameter analysis, Chin addresses confidence intervals related to how maximum likelihood estimates of parameters of interest would change with changes in nuisance parameters (i.e., parameters not being estimated). Chin expresses sensitivity to nuisance parameters as "the correction to the mean square errors of parameters of interest due to the changes in the nuisance parameters" (Chin, 1991). The concept of sensitivity to nuisance parameters, in this case, is applied to maximum likelihood estimates of the direction-of-arrival (DOA) of narrowband signals received by a sensor array. In light of this application, breadth of distribution analysis can be viewed as an extension of this concept with specific relation to quantification of user-interface usage.

E. SRI IDES STATISTICAL ANOMALY DETECTOR

Javitz and Valdes implement a statistical subsystem in a real-time intrusion-detection system that uses multivariate methods to profile normal behavior and identify deviations from expected behavior. Normal behavior is defined by the actions of individual users and groups of users over a period of time, so that the statistical subsystem adaptively learns to differentiate between normal and abnormal behavior in identifying potential intrusion. The statistical procedures used in determining anomalous behavior - which include means, covariances, and frequency tables - evaluate the total usage pattern for a particular individual or group by looking at the correlation between audit variables, as well as observing whether a particular audit variable is too high or too low. This study incorporates a variety of statistical methods that capture distributional information on individual users and groups of users, thus providing a set of statistical methods that may be used in breadth of distribution analysis of user interface usage.

F. STATISTICAL TESTING FOR CLEANROOM SOFTWARE ENGINEERING

In order to certify software quality, Whittaker and Poore use a stochastic model for the statistical certification testing phase of Cleanroom software engineering methodology. Two steps are involved in the statistical testing process:

- determine usage distribution in the intended environment by performing a top-down structural investigation of the specification document that establishes a set of usage states and defines an ordering on this set;
- use a statistical evaluation mechanism on the resulting sequences as they are applied to the software (Whittaker, 1992).

This study defines two standard software quality measures for determining the effect of an individual failure in the testing process: reliability and mean time between failure (MTBF). Reliability is the probability that a software run will not result in a failure, whereas MTBF is the mean recurrence time for the set of failure states in the chain (Whittaker, 1992). The combination of these central tendency measures with Markov chains produces a detailed analysis of expected usage patterns, which can be further enhanced through the use of breadth of distribution

analysis by providing a picture of the spread of these expected usage patterns.

G. EMPIRICAL VALIDATION OF USER-INTERFACE CHARACTERISTIC MEASUREMENTS

Since standards with regards to usability cannot measure all relevant features in a task-independent way, Rauterberg approaches quantitative measurement of user interface quality by first defining a means for describing user interfaces on a granularity level general enough to apply to various known interface types and detailed enough to preserve important interface characteristics. Different types of interfaces - such as graphical user interface (GUI), command language, multimedia, etc. - are distinguished and quantified by the general concept of "interaction points" (Rauterberg, 1996). Within an interactive system, the author distinguishes between application objects (e.g., a text document) and dialog objects (e.g., a window), and application functions (e.g., insertion section mark) and dialog functions (e.g., close window). A dialog context is defined by all available

functions and objects in the system state (Rauterberg, 1996). The quantitative measures the author uses are:

- functional feedback: the number of perceptible dialog functions per number of hidden dialog functions;
- interactive directness: the average length of all possible sequences of interactive operations from the top level dialog context down to the desired dialog context;
- application flexibility: the average number of hidden application function interaction points per dialog context;
- dialog flexibility: the average number of hidden application function interaction points per dialog context.

These measures enable a quantitative explanation for qualitatively observed differences between interfaces, thus providing a potential set of measures for use in our model. Breadth of distribution statistical methods would produce a detailed view of the range of values for these quantitative measures.

H. WWW SURVEY TOOL

The second WWW User Survey, conducted by Pitkow and Recker, used adaptive questions so that respondents were only presented with questions relative to their previous answers. The authors also used methods for tracking user responses across different sections of the survey, allowing more detailed analysis of survey responses (Pitkow, 1995). The results of this study show a widely disparate user base, which, in conjunction with the methods for adaptive questioning, provide a potential frame of reference for the design of our model.

I. WWW BROWSING STRATEGIES

Catledge and Pitkow conducted a study with the purpose of providing detailed information on actual WWW browsing strategies by using log file analysis on captured client-side user events, thus providing a set of concrete data for potential use in WWW design and usability guidelines. The results of this study allowed the authors to define separate user groups within the user population based on their usage characteristics, thus defining user groups with different sets of interests and needs. This information, along with

the methodology used in capturing client-side user events, provides a basis on which we can design an initial model for breadth of distribution analysis.

J. COMBINED LOG SYSTEM

The processing of numerous log files requires a tool for reading these files and sorting the results into useful sets of data. Beckett designs a combined log system, through which log file information can be automatically summarized into statistics on access types, operations, access times, data transfer size, etc. This is done by providing a combined format for the transfers, and converting each of the raw log files into this combined format. Instant access to the totals are available, split by access methods and summaries of users, sites and areas are also possible, although not saved by default. This tool is thus capable of capturing data which may be relevant to our study of breadth of distribution analysis.

III. COMPARISON MODEL

A. PREFACE

Through the application of appropriate statistical methods, breadth of distribution analysis provides a means of quantifying the degree of variance in a user population as a basis for analyzing a user-interface. The topological methods employed allow for the evaluation or characterization of the manner in which the differences among users of different interest and knowledge levels factor in the collected usage data. The selection and design of an interface whereby data can be easily gathered and analyzed on user behaviors will, in turn, facilitate the development of a set of recommendations for possible refinements to the interface in order to more closely match the needs of the user population.

B. SELECTING A MODEL: WORLD WIDE WEB (WWW) SITE

With the advent of Java as a major programming language, a perceived need for an on-line Java tutorial within the Naval Postgraduate School Computer Science

department led to the decision to use Java as a testbed for the model. One factor that affected this decision focused on the fact that the anticipated sample group was familiar with navigation through HTML documents, so that any difficulties in using the model result from the model itself and not the use of the browser.

Another factor that influenced the decision to use this tutorial as a testbed included the ease of implementation involved and the ease through which data could be collected on actual user behaviors. Data on web site usage is maintained by the system administrator through the use of log files, which record all accesses made to the particular web server. These log files include information such as:

- the machine name from which the file was accessed;
- a time stamp for when the file was accessed (date/time);
- the name of the file accessed;
- and the size of the file accessed.

The analysis of these log files allow the determination of the navigation patterns for each user, producing data that

provide information on the perceived usability for each particular user of the interface.

The navigation patterns of each user comes from analyzing the concrete data compiled on the user-interface interaction points within the model. The web site contains a hierarchical structure, such that each page has a hypertext link to the introduction section, the table of contents, the previous page, and the next page in sequence. In order to record from which page a user goes to reach another page, all of the hypertext links are aliased to the actual document the user wishes to access. For example, if the user wishes to go from the Table of Contents section to the first section of Chapter Two, the link on the Table of Contents page will read "toc_part2a.htm" (meaning "going from Table of Contents to Chapter Two, first section") instead of just "part2a.htm." This aliasing is done while building the model by using the UNIX line command "ln <alias> <file being aliased>." Since aliasing provides a means for capturing specific navigation patterns, data collection on the usability of this tutorial can yield potentially useful recommendations for future modifications and enhancements.

C. TOPOLOGICAL METHODS

In attempting to quantify the perceived usability of the interface, analysis of the log file entries for the tutorial focuses on the choices made by each individual user in navigating through the different pages. The particular usage characteristics studied in order to distinguish the different users include connectedness, order of traversal, and shape. These characteristics differ semantically and logically from those found in traditional graphs due to the dynamic and flexible nature of hypertext.

In traditional directed hierarchical graphs (known as trees), the hierarchy mandates that a "parent-child" relationship exists between the nodes in the tree. A directed graph contains edges between nodes that indicate direction; this graph may contain directed edges going both ways between two particular nodes. If a directed edge is between node x and y, then x is the parent of y and, consequently, y is the child of x. This structure also requires the existence of a distinct root node, the only node with no parent and the one from which all other nodes branch. With the exception of the root node (since the root is the beginning of the tree), each node within the tree has a distinct parent node and may have anywhere from zero to n

child nodes, where n denotes the degree of the tree (n-ary). For example, a tree with n = 2 (a binary tree) will have nodes with either zero, one or two children. A node with no children is called a leaf node.

In a traditional directed graph, each node may only be visited using the edges defined within the tree. However, with hypertext a user has the ability to visit different nodes either through the pre-defined links between the different nodes (pages) or by simply typing in the name of the desired page, thus circumventing whatever hierarchical structure previously defined by the architect of the interface. The testbed for the Java tutorial in this model incorporates a logical hierarchy in the relation of topic areas (chapters, sections, sub-sections), but the links between these utilize a structure that allows for greater flexibility in navigation than allowed in traditional directed graphs. This flexibility employs the following design decisions:

- All chapters, sections, and sub-sections are accessible from the Table of Contents page, and each of these pages contains a link back to the Table of Contents page.

- Each page, in addition to the Table of Contents link, contains a link back to the Introduction page.

- Each page contains a link to the previous page and next page in the logical flow of the tutorial, so that the user may navigate through the tutorial as if reading a textbook on Java. The only exceptions to this rule are the first page following the Table of Contents (since the previous page is the Table of Contents, there is no previous page link), and the last page of the tutorial (no next page link).

Figure 1 (included at the end of this section) shows an illustration of the tutorial structure.

The issue of connectedness, in relation to hypertext documents, differs from the traditional concept of connectedness both semantically and logically in that it deals with examining the linkage between two consecutive WWW pages visited in the log file, since a user can get to a page either by linking directly from the current page, linking from a previous page, or simply typing in the URL of the desired page. In the case of by linking to a page directly from the current page, the link that registers in the log file is the dynamic link which identifies both the current page and the destination page. For example,

"intro_toc.htm" is a dynamic link for the Table of Contents page on the Introduction page, denoting the path traveled as "from the Introduction to the Table of Contents."

In the case of linking to a page from a previous page, the examination of two consecutive log file entries fail to show a logical progression. This occurs when the user uses the "Back" button on the WWW browser to return to an earlier page, then uses a link from this page. For example, if the log file has two consecutive entries that read "toc_part2d.htm" and "toc_part2e.htm," this shows that the user visited the page containing Part 4, Section D of the tutorial by using the hypertext link on the Table of Contents page, then used the "Back" button to return to the Table of Contents page, finally using the hypertext link on the Table of Contents page to visit the page containing Part 4, Section E.

In the case of linking to a page by simply typing in the URL of the desired page, the link that registers in the log file is likely the actual HTML file for that page, not one of the dynamic links on another page. The possibility exists that a user decides to link to another page by typing in the name of a dynamic link to that page, but this case is highly unlikely, since a user who types in the URL for a

specific page knows which section he wants and has no concern for documenting the page from which he arrives at the desired page. Whether a user arrived at a page by typing in the URL or not can be determined from the log file entry, which will contain either a dynamic link (such as "toc_part2d.htm") or the name of the actual HTML document (such as, in this particular case, "part2d.htm").

In order to define the degree of connectedness in hypertext in quantifiable terms, the need exists to measure the extent to which two consecutive log entries relate to each other. The degree to which the visitation of two consecutive pages is predicted by the architect of the interface, as well as the extent to which a logical relationship exists between the two pages, defines three general degrees of connectedness: zero, partial, and maximum. A high degree of connectedness in the log files would indicate that the interface meets the needs of the user population, whereas low degree of connectedness between entries would reflect a need to re-structure the interface to improve the ability for user to navigate through the interface logically to meet their needs.

Zero connectedness exists when the visitation of two consecutive HTML documents is neither logical (i.e., no

definable relationship between the content or location of the two files) nor predicted (the architect has no pre-defined link between the two documents). Zero connectedness occurs when the user visits a page by typing in the URL for the desired page, and no discernible logical relationship exists between the two pages.

Maximum connectedness exists when the visitation of two consecutive HTML documents is both logical (i.e., a clearly defined relationship exists between the content and/or location of the two files) and predicted (the user went from one document to the next using a link previously defined by the architect of the interface). If the log files reflects a prevalence of maximum connectedness between entries, the architect has succeeded in predicting user behavior with regards to the interface.

Partial connectedness exists when the visitation of two consecutive HTML documents is either logical or predicted, but not both. Partial connectedness thus may occur either when the user visits a page by using a link from the Table of Contents page (the design of the tutorial is such that every page may be reached from the Table of Contents), or the user types in the URL for the desired page and a logical

relationship exists between the desired page and the previous page.

The order of traversal for a particular user in navigating through the tutorial provides data on the directional nature of searches the user employs in obtaining the desired information from the tutorial. In general, the characterization of the usage patterns for a particular user may define a depth-first search or a breadth-first search strategy. Because of the size of the tutorial and the dynamic nature of hypertext, combinations of these two basic strategies may exist. For example, a user may conduct a breadth-first search across the main section headings, which may be done using consecutive links from the Table of Contents page. A user may also perform a depth-first search on a particular topic area, either using the pre-defined links on each page or by visiting the pages in a logical manner from the Table of Contents page. The fact that hypertext allows visitation of nodes (pages) that don't follow the links previously defined by the interface designer may result in a random pattern of traversal fora particular user. By analyzing the log file entries for each specific user, the identification of these strategies

provide a means for identifying the interests and needs of that particular user.

The issue of shape, in relation to hypertext documents, looks at the connectedness of the pages visited, the order of traversal in navigating through the given structure, and the time spent on each page. The shape of an individual's usage defines the scope of a user's interests and needs with relation to the interface by indicating a characterization of the user's methods of finding the desired information, and what actions the user takes once the information is found. For example, one user may initially incorporate a breadth-first search strategy within a particular topic area until finding the specifically desired page, then spending an extended period of time on that particular page, whereas another user may have a very specific knowledge of where to find the desired information, then upon arriving at the appropriate page, spends a brief period of time on that page before leaving the interface. Time spent on a particular page may reflect the level of interest in the page, the level of difficulty in understanding the content, or the actual size of the page. The examination of these issues based on data analyzed on actual usage provides an

indication of the interests and needs of the user population.

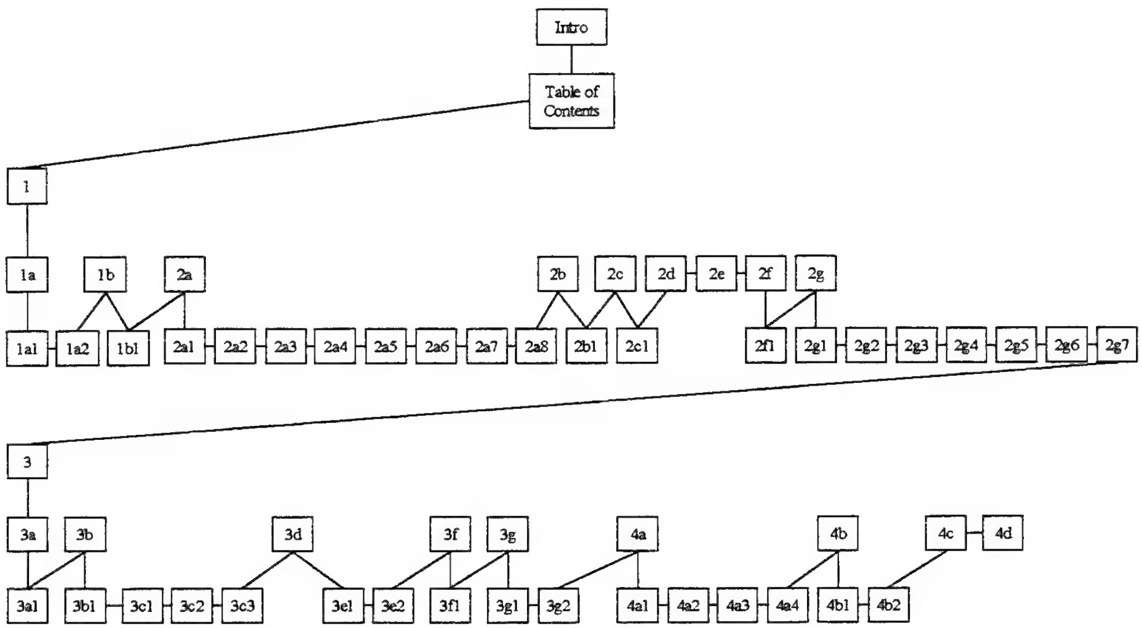

Figure 1. Page Structure of Java Tutorial, with Associated Links. Each line represents a link in both directions. In addition, all pages contain links to the Introduction and the Table of Contents, and the Table of Contents contains links to every page.

D. SUMMARY

Through log file analysis and the application of appropriate statistical methods in examining usage data related to this tutorial, breadth of distribution analysis provides a means of quantifying the degree of variance in

observed user behaviors as a basis for analyzing the effectiveness of the user-interface in meeting the needs of the user population. By measuring the attributes of connectedness, size, and traversal order in relation to these hypertext documents, the data collected should enable the characterization of the differences among users of different interest and knowledge levels with regards to the interface in question. The selection and design of a WWW Java tutorial facilitates the collection and analysis of data and, thus, allows for the development of a set of recommendations for possible refinements to the interface in order to more closely match the needs of the user population.

IV. DESIGN AND IMPLEMENTATION

A. PREFACE

One approach to user-interface optimization utilizes statistical hypothesis testing on specific user behaviors in conjunction with the interface. Hypothesis testing tests an assumption (the hypothesis) we have about the targeted user population against a sample from the population. The result of a significance test is a probability that is attached to a descriptive statistic calculated from a sample. Prior to examining the data collected on the usage of the tutorial, we postulated several hypotheses to characterize the predicted use of the interface based on observations and perceptions of the tutorial structure and the user population.

B. HYPOTHESIS FORMULATION

One hypothesis of interest involves what percentage of the users will make use of the built-in links in their navigation patterns, and what percentage will use the browser's navigation buttons. We estimate that

approximately two-thirds of the users will follow the built-in links to find the desired information, instead of using the WWW browser's "Back" button and the Table of Contents links. The use of this figure comes from the fact that, with the exception of the Introduction and Table of Contents pages, and the first and last pages of the actual tutorial, the user has four link options: Introduction, Table of Contents, "Next" and "Previous." Since the Introduction page only provides a link to the Table of Contents, users will likely never link back to the Introduction page from the current page. Of the three remaining links, only two - "Next" and "Previous" - provide the user with the ability to follow the tutorial as if reading through a textbook. Consequently, this hypothesis is based on the notion that the users will prefer to follow a navigation pattern that closely resembles their natural reading patterns, thus making the assumption that users will typically forego using the browser's navigational buttons, and providing a measure for the degree of connectedness of the tutorial interface.

Another issue of interest is the degree of variance between the shape and order of traversal associated with the navigation patterns of different users within the test sample. We hypothesize that by observing these navigation patterns that there exists specific, identifiable user sub-

groups within the sample. The existence of such sub-groups would imply the existence of noticeably distinct groups of users within the user population, thus necessitating the consideration of different sets of needs and interests within this population in adding future enhancements to the interface.

C. MEASUREMENT CRITERIA

The population in this particular case involved the student population of the Naval Postgraduate School's Computer Science, Electrical & Computer Engineering, Information Technology Management, and Information Warfare curricula. Each of these curricula require varying degrees of computer programming knowledge as part of the course matrix for every student. With the advent of Java and the growth in popularity of the World Wide Web, a high demand for opportunities for learning Java as a programming language has developed, whether these opportunities come through classroom instruction, directed study, or other means. The majority of the students in these curricula do not own bachelor's degrees in their respective field of study, although the majority have taken at least one

introductory programming class (usually Pascal, BASIC, or FORTRAN) at the college level. Of the students with Bachelor's degrees in their respective fields, most have no prior experience programming in Java.

The selection of statistical methods for this model focused on the applicability of various statistics towards user-interface design, in order to maximize the relevance of the data collected on user behaviors. The process of choosing these methods also examined what statistics would provide a quantifiable distinction between the needs and interests of particular user sub-groups within the observed sample. Using statistical inference allows the characterization of a population on the basis of the observed sample, and provides a means for determining how well that sample fits our preconceptions of the population from which the sample was drawn.

In testing the second hypothesis, Analysis of Variance (ANOVA) provides a means for testing the difference between two or more means by examining the ratio of variability (variance) between two conditions and the variability within each condition. Application of ANOVA to usage data collected for individual users should provide a means for subcategorizing usage patterns amongst users within the

sample. Thus, when the variability between the two groups is much greater than the variability within each group, then we will conclude that distinct user subgroups exist within the sample.

D. RESULTS OF DATA COLLECTION

The data collected showed various individual usage strategies, differing in shape, connectedness and order of traversal. The majority of the users followed the built-in links while navigating through the tutorial, implying a high degree of connectedness in the tutorial design. Some users followed a more centralized approach whereby they used the Table of Contents page as a central hub from which they would jump to the particular topic of interest. Within this subgroup, the majority used the built-in links back to the Table of Contents, while others used the browser's Back button to return to the Table of Contents. As expected, the pages visited most often were the Introduction page and the Table of Contents page, since the URL advertised to students for the tutorial was the URL for the Introduction page, and the Introduction page linked only to the Table of Contents page. However, after examining which visits to the

Introduction and Table of Contents pages resulted from the initial arrival to the tutorial, more than 50 per cent of the total Table Of Contents page visits occurred as part of the navigation process through the tutorial. These results suggest that a percentage of the users in the sample utilized the built-in Table of Contents links in navigating through the tutorial.

In order to determine what percentage of users followed the built-in "Next" and "Previous" links versus what percentage used the Table of Contents as a central node for navigation, the ratio of "Next" and "Previous" links to the total number of page visits must be determined, excluding the initial Introduction and Table of Contents page links. This ratio for each user in the sample is calculated as:

$$\frac{\text{\# "Next" and "Previous" links}}{\text{\# total page visits - initial links}}$$

Let X represent the number of "Next" and "Previous" links used, Y represent the number of total page visits excluding the initial visits to the Introduction and Table of Contents pages, and R represent the calculated ratio between X and Y. The values for X, Y, and R for each user in the sample are shown below in Table 1:

USERS	X	Y	R
User A	7	8	0.875
User B	4	11	0.364
User C	33	57	0.579
User D	0	2	0.000
User E	17	25	0.680
User F	5	6	0.833
User G	0	2	0.000
User H	8	10	0.800
User I	59	62	0.952
User J	9	9	1.000
User K	0	35	0.000
User L	64	64	1.000
User M	58	58	1.000
User N	26	28	0.929
User O	2	4	0.500
User P	1	4	0.250
User Q	3	3	1.000
TOTAL	296	388	0.763

Table 1. Connectedness Ratio Calculations

Since the hypothesis postulated that this ratio would be greater than two-thirds, the application of a one-tailed small-sample inference test on the difference between the test value and the observed value will provide information on the validity of the observed value. The calculations of this test are shown below in Table 2:

$x = 0.763$	$t = (x - \mu_0)/(s/n^{1/2})$
$\mu_0 = 0.667$	$t = (0.763 - 0.667)/(0.402/4)$
$s = 0.402$	$t = 0.955$
$n^{1/2} = 4$	$t_\alpha = 1.337$, $\alpha = 0.10$ (16 degrees of freedom)

Table 2. One-Tailed, Small-Sample Inference Test Calculations

Since the observed value for t is less than t_α, the null hypothesis cannot be rejected; in other words, the results fail to support the hypothesis that more than two-thirds of the page visits made by users in the sample utilized the built-in "Next" and "Previous" links. These results show how breadth of distribution statistical methods can take a measure of central tendency and determine the significance of the measured value.

In examining the second hypothesis, the use of graphs to provide a pictorial representation of a specific user's navigation pattern, along with data on the average time spent per page and the number of pages visited, helped to initially break out potential user sub-groups. The three graphs that follow serve as examples to illustrate the navigation patterns of three users from the sample (Figures 2 through 4). In Figure 2, User B follows an approach whereby a topic of interest is found near the beginning of usage, then the user follows the tutorial by browsing through a few other topics until exiting the tutorial. The breadth of this user's browsing strategy tends to focus on a particular area of interest rather than across the scope of the tutorial.

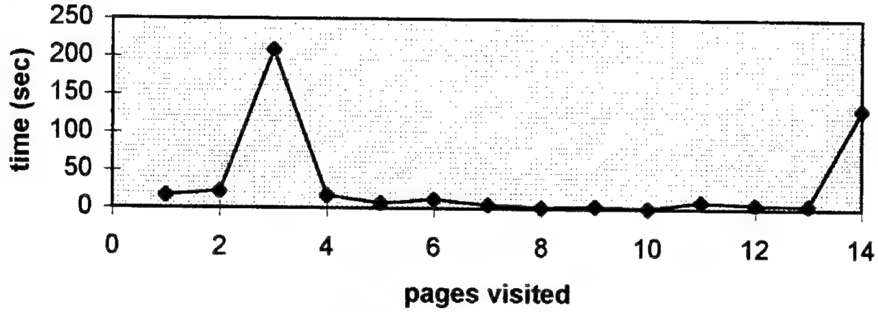

Figure 2. Usage Graph for User B

In Figure 3, User I follows an approach characterized by more extensive browsing than User B. As User I finds a topic of interest, he spends additional time on that particular page, then continues browsing until he finds another topic of interest or arrives at the end of the tutorial. The breadth of this user's browsing strategy differs from that of User B in that it ranges across the scope of the tutorial rather than focusing on specific topics or a particular area of interest.

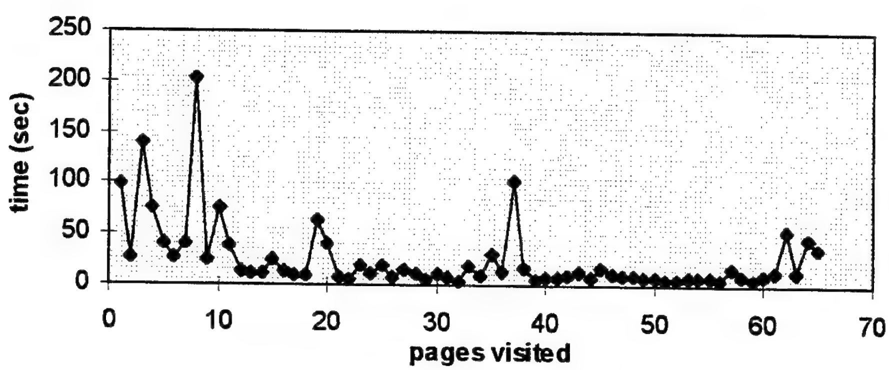

Figure 3: Usage Graph for User I

In Figure 4, User P follows an approach characterized by few pages visited, but on average a substantial amount of time spent on each page. As User P walks through the tutorial, he spends extra time on each page, then continues to the next topic provided by the built-in link. The breadth of this user's browsing strategy is thus similar to that of User B, but differs from User B in shape.

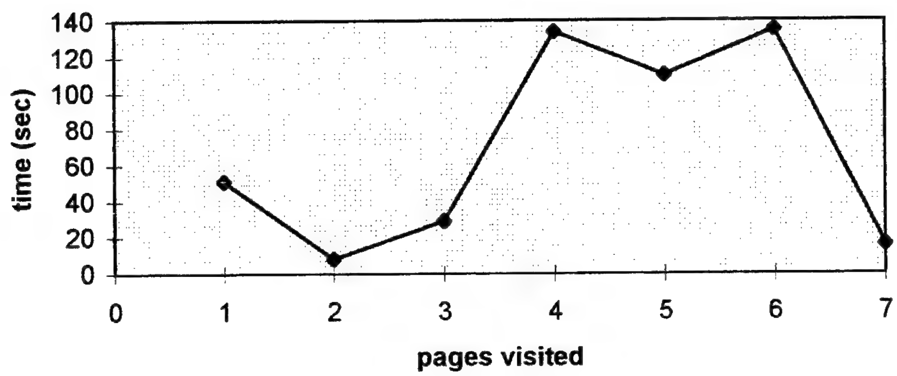

Figure 4: Usage Graph for User P

After examining the usage data from the sample, the users were divided into four groups of similar usage characteristics, through which application of ANOVA would test the relevance of the sub-group selection. Calculations for the average and variance of each sub-group, along with the number of users in each, are listed in Table 3 below:

	AVG	VAR	n_k
Group 1	93.14	478.70	4
Group 2	53.18	92.82	4
Group 3	30.34	8.57	5
Group 4	21.07	13.91	4

Table 3. Average, Variance & Size for User Sub-Groups

The performance of the ANOVA test compares each group with the other three, with the results shown below in Table 4:

$N = 17$

$T = 821.31707$

$T^2/N = 39680.10$

$n_1 = 4$

$T_1 = 372.5767 \quad T_2 = 212.7222 \quad T_3 = 151.723 \quad T_4 = 84.29517$

$T_1^2 = 138813.40 \quad T_2^2 = 45250.73 \quad T_3^2 = 23019.87 \quad T_4^2 = 7105.676$

$T_1^2/N = 33453.35$

$T_2^2/N = 11312.68$

$T_3^2/N = 4603.97$

$T_4^2/N = 1776.42$

$\Sigma\ \Sigma\ X^2 = 54781.00$

$SST = \Sigma\ \Sigma\ X^2 - T^2/N = 54781.00 - 39680.10 = 15100.9$

$SSA = \Sigma\ T_k^2/N - T^2/N = 51146.42 - 39680.10 = 11466.32$

$SSE = SST - SSA = 15100.9 - 11466.32 = 3634.58$

$MSE = \sigma^2 \text{ (pooled)} = SSE/(n_1 + n_2 + n_3 + n_4 - 4) = 279.58$

$MSA = SSA/(K - 1) = 11466.32/3 = 3822.11$

$F = MSA/MSE = 3822.11/279.58 = 13.67$

Table 4. ANOVA Ratio Calculations for User Sub-Groups

Thus, the ratios generated by the ANOVA test suggest the existence of four distinct sub-groups within the user sample, from which conclusions may be derived for the population.

E. SUMMARY/CONCLUSION

After determining a set of hypotheses for examination and selecting appropriate statistical methods, the compilation of concrete data on the various interaction points of the user-interface provided a basis for determining the breadth of distribution. By applying these statistical methods to the data collected, distinct characterizations of specific user sub-groups within the sample provided a means for identifying potential areas for enhancement in order to more closely meet the needs and interests of the user population. Future measurement of any design modifications implemented as a result of this study can provide more information as to how well the modifications have addressed the perceived needs of the user population.

V. SUMMARY AND CONCLUSIONS

A. SUMMARY

The evaluation of user interfaces using centrally-based measures may ignore large populations of "non-average" users by failing to address the distribution of users within the population. These centrally-based measures provide a value intermediate to the set of values in question, along with an average degree of variation from that intermediate quantity. Breadth of distribution analysis employs measures of extent of a distribution as the basis for analyzing a user interface. This approach to user interface analysis quantifies the degree of variance for heterogeneous groups of users by examining the spread of values over a graph of the interface usage. Thus, breadth of distribution analysis breaks out the sub-groups that exist within a user population as represented pictorially by separate but distinct curves in the graph.

After surveying the current literature on the issue of measuring user interface characteristics towards the goal of improved usability, this thesis selected the statistical methods for use in breadth of distribution analysis. Upon determining which statistical methods to use, the design of

a testbed for hypothesis testing implemented a World Wide Web site for use in data capture. Log file analysis enabled the collection of concrete data on user behaviors associated with the user interface. The implementation of the selected statistical methods on the log file analysis results allowed the formulation of conclusions regarding user interests and needs in certain topic areas within the interface, as well as recommendations for possible refinements to the interface in order to properly address the needs of the user population.

Another value of this analysis method lies in its ability to show significance of results. Differences that intuitively appear interesting may be shown to be insignificant. This thesis demonstrated identification of significance analysis of web page linkage.

B. RECOMMENDATIONS

As information technologies continue to evolve, the increasing accessibility of these technologies will continue to attract users of different experience levels, skill levels, computer usage frequency, and requirements for as well as interests in information technologies. Their

enduring success depends on their ability to adjust to the disparate needs and interests of the user population. The development of intelligent user-centered interfaces provides for intuitive HCI, which in turn contributes to greater initial and long-term interest in information technology. Breadth of distribution analysis provides further illumination on the needs and interests of the user population, thus enhancing the design of user interfaces responsive to various groups of users.

Within the field of user interface analysis, several user interface design methodologies exist, each with its own set of measurements. Since studies using each of these methodologies have historically focused on central tendency measures, the application of breadth of distribution analysis towards these measurements may yield useful information towards usability recommendations.

With the continued growth of the World Wide Web, the size and diversity of the user population will continue to expand. The development of intelligent user-centered hypertext interfaces will contribute to greater initial and long-term interest in the Web. Since breadth of distribution analysis provides a characterization of distinct sub-groups within the user population, the

application of breadth of distribution analysis will allow Web publishers to more closely identify the needs and interests of the user population.

As consumers expand their use of software and the World Wide Web to complete tasks and purchase other products and services, the ability of a software product or Web site to meet the needs of targeted consumer groups will affect its ability to attract and maintain user interest. Thus, a keen understanding of the present and future needs and interests of the user population is necessary for the design of user-interfaces responsive to user needs and interests. Breadth of distribution analysis provides a means for marketing researchers and interface developers to produce an interface that acts as an effective vehicle for communication with the user.

C. FUTURE WORK

While this study addressed the applicability of breadth of distribution measures towards user interface analysis, several areas of research with regards to user interface analysis may be pursued from this thesis. This application of breadth of distribution analysis incorporated a specific

set of measures that were determined to provide the information necessary to examine the hypotheses made on the usage of the interface. In addition, the characteristics of the selected interface included only a subset of the potential features associated with a user interface.

One area in which the study of breadth of distribution analysis may prove useful is in the development of randomly generated interfaces. The concept behind randomly generated interfaces focuses on the idea of changing the interface dynamically in response to user input, so that the user behavior in conjunction with the interface directly affects the presentation of the interface. The identification of the needs and interests of different user sub-groups can provide a means for determining how an interface will automatically adapt to behaviors exhibited by a particular user. In addition to providing a means for measuring how well an interface meets the needs and interests of the user sub-groups within the population, these measures may provide a metric for determining how accurately the adaptive interfaces update in accordance with observed user behaviors.

Further research with breadth of distribution analysis may also examine the effect of various graphical features on

the perceived usability of an interface. In order to minimize download time, the interface in this study only made use of text and hypertext links. Other studies may explore the ability to utilize breadth of distribution methods in measuring the effect of color, graphical icons, applets, and other graphical items on the usability of an interface. Breadth of distribution measures may provide an effect metric in determining the usability of various interface elements and designs.

Continued research in the utilization of breadth of distribution methods may yield useful information in the continuous process of meeting the needs and interests of the user population. This study, like any first study, addressed only some of the issues and potential applicability of breadth of distribution analysis, which merit further investigation and study. The analysis of the applicability of various breadth of distribution measures on different user interface issues and features should yield information useful to the continued development of user interface analysis techniques.

LIST OF REFERENCES

Balasubramanian, V., and Turoff, Murray, *A Systematic Approach to User Interface Design for Hypertext Systems,* Proceedings of the 28^{th} Annual Hawaii International Conference on System Sciences, 1995.

Bass, Len and Coutaz, Joelle, *Developing Software for the User Interface,* Addison-Wesley Publishing Company, Inc., 1991.

Catledge, Lara D. and Pitkow, James E., *Characterizing Browsing Strategies in the World-Wide Web,* Computer Networks and ISDN Systems, Vol. 27, 1995.

Chin, Daniel C., *On the Connection Between Maximum Likelihood Sensitivity Analysis and Nuisance Parameter Analysis,* IEEE Transactions on Aerospace and Electronic Systems, Vol. 28, No. 3, July 1992.

Dumas, Joseph S., *Designing User Interfaces for Software,* Prentice-Hall, Inc., 1988.

Henckel, Ramon E., *Tests of Significance,* Sage Publications, Inc., 1976.

Malone, Thomas B., *Human Factors and Human Error,* Proceedings of the Human Factors Society 34th Annual Meeting, Vol. 1, 1990.

Pitkow, James E. and Recker, Margaret M., *Using the Web as a Survey Tool: results from the second WWW user survey,* Computer Networks and ISDN Systems, Vol. 27, 1995.

Rauterberg, Matthias, *An Empirical Validation of Four Different Measures to Quantify User Interface Characteristics Based on a General Descriptive Concept for Interaction Points,* Proceedings of IEEE Symposium and Workshop on Engineering of Computer-Based Systems, IEEE Computer Society Press, 1996.

Rencken, W.D. and Durrant-Whyte, H.F., *A Quantitative Model for Adaptive Task Allocation in Human-Computer Interfaces,* IEEE Transactions on Systems, Man, and Cybernetics, Vol. 23, No. 4, July/August 1993.

Whittaker, James A. and Poore, J.H., *Statistical Testing for Cleanroom Software Engineering,* Proceedings of the Annual Hawaii International Conference on System Sciences, January 1992.

BIBLIOGRAPHY

Balasubramanian, V., and Turoff, Murray, *A Systematic Approach to User Interface Design for Hypertext Systems,* Proceedings of the 28^{th} Annual Hawaii International Conference on System Sciences, 1995.

Bass, Len and Coutaz, Joelle, *Developing Software for the User Interface,* Addison-Wesley Publishing Company, Inc., 1991.

Catledge, Lara D. and Pitkow, James E., *Characterizing Browsing Strategies in the World-Wide Web,* Computer Networks and ISDN Systems, Vol. 27, 1995.

Chin, Daniel C., *On the Connection Between Maximum Likelihood Sensitivity Analysis and Nuisance Parameter Analysis,* IEEE Transactions on Aerospace and Electronic Systems, Vol. 28, No. 3, July 1992.

Dumas, Joseph S., *Designing User Interfaces for Software,* Prentice-Hall, Inc., 1988.

Henckel, Ramon E., *Tests of Significance,* Sage Publications, Inc., 1976.

Javitz, Harold S. and Valdes, Alfonso, *The SRI IDES Statistical Anomaly Detector,* IEEE Symposium on Research in Security and Privacy, 1991.

Malone, Thomas B., *Human Factors and Human Error,* Proceedings of the Human Factors Society 34th Annual Meeting, Vol. 1, 1990.

Pitkow, James E. and Recker, Margaret M., *Using the Web as a Survey Tool: results from the second WWW user survey,* Computer Networks and ISDN Systems, Vol. 27, 1995.

Rauterberg, Matthias, *An Empirical Validation of Four Different Measures to Quantify User Interface Characteristics Based on a General Descriptive Concept for Interaction Points,* Proceedings of IEEE Symposium and Workshop on Engineering of Computer-Based Systems, IEEE Computer Society Press, 1996.

Rencken, W.D. and Durrant-Whyte, H.F., *A Quantitative Model for Adaptive Task Allocation in Human-Computer Interfaces,* IEEE Transactions on Systems, Man, and Cybernetics, Vol. 23, No. 4, July/August 1993.

Whittaker, James A. and Poore, J.H., *Statistical Testing for Cleanroom Software Engineering,* Proceedings of the Annual Hawaii International Conference on System Sciences, January 1992.

DISTRIBUTION LIST

1. Defense Technical Information Center 2
   8725 John J. Kingman Rd., STE 0944
   Ft. Belvoir, Virginia 22060-6218

2. Dudley Knox Library 2
   Naval Postgraduate School
   411 Dyer Rd.
   Monterey, California 93943-5101

3. Chairman, Code CS 2
   Computer Science Department
   Naval Postgraduate School
   Monterey, CA 93943-5000

4. Timothy Shimeall 2
   Code CS/SM
   Computer Science Department
   Naval Postgraduate School
   Monterey, CA 93945-5100

5. John Daley 1
   Code CS/DA
   Computer Science Department
   Naval Postgraduate School
   Monterey, CA 93945-5100

6. John Arquilla 1
   Code IW/AR
   Information Warfare Academic Group
   Naval Postgraduate School
   Monterey, CA 93945-5100

7. Superintendent 1
   (Code EC/LT)
   Naval Postgraduate School
   Monterey, CA 93943-5121

8. Commanding Officer 2
   (Attn: Code 30, CDR Zellmann)
   Naval Information Warfare Activity
   9800 Savage Rd.
   Ft. Meade, MD 20755-6000

9. ECJ6-NP 1
HQ USEUCOM
Unit 30400 Box 1000
APO AE 09128

10. Dr. Luis A. Reinoso 1
2430 E. 41st St.
Tulsa, OK 74105